

Review of Master's Thesis

Student: Horký Michal, Bc.
Title: Beat Grep with Counters, Challenge (id 22461)
Reviewer: Češka Milan, RNDr., Ph.D., DITS FIT BUT

- 1. Assignment complexity** **less demanding assignment**
Zadání považuji za mírně obtížnější, jelikož se jedná o re-implementaci pokročilé metody pro vyhledávání regulárních výrazů. Metoda je založena na využití čítačových automatů a jejich determinizace a tudíž student musel nastudovat a pochopit pokročilé partie teorie automatů.
- 2. Completeness of assignment requirements** **assignment fulfilled**
Zadání bylo řádně splněno. Jedinou výhradu mám k experimentálnímu vyhodnocení, které mohlo být provedeno do větší hloubky (viz dále).
- 3. Length of technical report** **in usual extent**
Rozsah technické zprávy je v obvyklém rozmezí. Jednotlivé části jsou informačně přínosné. I když je velká část práce věnována popisu základních i pokročilých technik pro vyhledávání regulárních výrazů, považuji tyto části za velmi přínosné. Popis implementační části je dostatečný. Jedinou výhradu mám k rozsahu části popisující experimentální vyhodnocení.
- 4. Presentation level of technical report** **85 p. (B)**
Prezentace technické zprávy včetně logické struktury a návaznosti kapitol je na velmi dobré úrovni. I když DP prezentuje technicky komplikované principy, je dobře pochopitelná. K prezentaci mám následující výhrady. Ačkoliv je úvod práce velice pěkně zpracovaný, postrádá odstavec, který by shrnul dosažené výsledky. Obrázky 4.8 - 4.12 jsou převzaté (což je řádně uvedeno), ale jejich kvalita není dostatečná - jedná se o jednodušší obrázky a tak je mohl student překreslit. Má hlavní výhrada souvisí s již zmiňovanou experimentální částí. Zde bych očekával detailnější interpretaci výsledků včetně rozboru, který by ukázal, na jakých typech regulárních výrazů je implementovaná technika efektivnější a proč. DP obsahuje pouze shrnutí výsledků přes rozsáhlou sadu regulárních výrazů.
- 5. Formal aspects of technical report** **90 p. (A)**
Formální úprava práce je až na pár drobností (např. výše zmíněné obrázky, či nevhodně zalomená nadpisy kapitol) na vynikající úrovni. Jazyková úroveň angličtiny je taktéž velmi dobrá.
- 6. Literature usage** **90 p. (A)**
Student demonstroval dostatečné pochopení problematiky a orientaci v existující literatuře. Existující přístupy jsou velmi dobře pokryty a prezentovány. Citační pravidla jsou dodržena a jednotlivé citace jsou v souladu s citačními zvyklostmi.
- 7. Implementation results** **85 p. (B)**
Na základě demonstrace vytvořeného nástroje mohu konstatovat, že nástroj je funkční. Během demonstrace student prokázal orientaci a pochopení hlavních částí kódu. Tyto části jsou rovněž rozumně komentovány. Validace a testování je modulární a dobře zpracovaná. Výhrady mám k experimentální části - zde bych očekával nástroje (skripty), které dovolí lepší interpretaci a rozbor výsledků.
- 8. Utilizability of results**
Text práce je spíše kompilačního charakteru, ale realizační výstup je rozhodně přínosný. Jak je demonstrováno v experimentální části, vytvořený nástroj je zásadně rychlejší než původní implementace vyhledávání regulárních výrazů pomocí čítačových automatů. Tudíž, na jisté třídě regulárních výrazů nástroj poskytuje nejrychlejší známé vyhledávání. Celková využitelnost DP je bohužel limitována slabším experimentálním vyhodnocením.
- 9. Questions for defence**
 1. Můžete charakterizovat třídu (a uvést konkrétní příklad) regulárních výrazů (a případně textů pro vyhledávání), na které Váš nástroj překonává existující nástroje (grep, RE2, CA)?
 2. Nástroje jako jsou grep a RE2 těží z pokročilé optimalizace. Domníváte se, že výkonnost Vašeho nástroje lze značně vylepšit dalšími optimalizacemi (případně jakými)?
- 10. Total assessment** **85 p. very good (B)**
Práci celkově hodnotím velice kladně a to zejména kvůli obtížnějšímu zadání, kvalitní technické zprávě a dosaženým realizačním výsledkům. Jak je uvedeno výše, hlavní nedostatek práce vidím v experimentálním vyhodnocení, které nejde dostatečně do hloubky a není tak možné dosažené výsledky dobře interpretovat. Celkově navrhuji práci hodnotit známkou B.

In Brno 30 May 2021

Češka Milan, RNDr., Ph.D.
reviewer